

Supplementary Material

NPASS: Natural product activity and species source database for natural product research, discovery and tool development

Xian Zeng^{1,2}, Peng Zhang², Weidong He², Chu Qin², Shangying Chen², Lin Tao^{2,5}, Yali Wang², Ying Tan¹, Dan Gao¹, Bohua Wang^{3,4}, Zhe Chen⁵, Weiping Chen^{4*}, Yu Yang Jiang^{1*}, Yu Zong Chen^{1,2*}

¹ Breeding Base-Shenzhen Key Laboratory of Chemical Biology, the Graduate School at Shenzhen, Tsinghua University, Shenzhen Kivita Innovative Drug Discovery Institute, Shenzhen 518055, P. R. China.

² Bioinformatics and Drug Design group, Department of Pharmacy, National University of Singapore, Singapore 117543, Singapore.

³ Key Lab of Agricultural Products Processing and Quality Control of Nanchang City, Jiangxi Agricultural University, Nanchang, 330045, P. R. China.

⁴ College of Life and Environmental Sciences, Collaborative Innovation Center for Efficient and Health Production of Fisheries in Hunan Province, Hunan University of Arts and Science, Changde, Hunan, 415000, P. R. China.

⁵ Zhejiang Key Laboratory of Gastro-intestinal Pathophysiology, Zhejiang Hospital of Traditional Chinese Medicine, Zhejiang Chinese Medical University, School of Medicine, Hangzhou Normal University, Hangzhou 310006, R. P. China.

* To whom correspondence should be addressed. Y.Z. Chen Tel: +65 6516 6877; Fax: +65 6774 6756; Email: phacyz@nus.edu.sg.

Correspondence may also be addressed to Y.Y. Jiang Tel: +86 755 2603 6430; Fax: +86 755 2603 6430; Email: jiangyy@sz.tsinghua.edu.cn and W.P. Chen Tel.: +86 791 8381 3420. Fax: +86 791 8381 3655. E-mail: iaochen@163.com.

Table of content

1. Supplementary Table S1. Distribution of NPs in each chemical superclass.
2. Supplementary Table S2. Distribution of NPs in each source kingdom or super-kingdom.
3. Supplementary Table S3. Data sources of source species of NPs.

















SuperClass	Total NPs	No. of active NPs *	No. of potent NPs*
Lipids and lipid-like molecules	11,376	2,265	780
Phenylpropanoids and polyketides	5,093	1,074	380
Organoheterocyclic compounds	4,713	1,102	440
Benzenoids	2,801	753	287
Organic acids and derivatives	1,983	378	259
Organic oxygen compounds	1,880	295	114
Alkaloids and derivatives	1,143	294	156
Lignans, neolignans and related compounds	783	159	95
Organic nitrogen compounds	123	39	23
Nucleosides, nucleotides, and analogues	122	51	35
Organohalogen compounds	58	12	5
Organosulfur compounds	52	17	9
Hydrocarbons	50	6	2
Organic Polymers	33	11	1
Hydrocarbon derivatives	23	10	10
Organic 1,3-dipolar compounds	9	2	0
Organometallic compounds	4	1	0
Acetylides	3	0	0

Supplementary Table S1. Distribution of NPs in each chemical superclass. NPs are classified based on ClassyFire chemical taxonomy. Gradient red color indicates relative number of NPs in each superclass. Green bars indicate percentage of active/potent NPs in each superclass (full-filled bar in the cell is 100%).

*Given biological activities of NPs are measured against multiple types of targets (e.g. proteins, microorganisms, cell lines), potent NPs are tentatively defined as that a NP has activity values (IC50/EC50/Ki/MIC50/GI50/AC50/Potency) of ≤ 500 nM against any protein targets or $\leq 1\mu\text{M}$ against any microorganism or cell line targets. While corresponding cutoffs of active NPs are tentatively defined as $500\text{ nM} < \text{activity values} \leq 5\mu\text{M}$ or $1\mu\text{M} < \text{activity values} \leq 10\mu\text{M}$ against each type of targets, respectively.

Superkingdom/Kingdom of Source Species	No. of source species	Total NPs	No. of active NPs *	No. of potent NPs*
Viridiplantae	16,431	21,338	4,780	1,736
Metazoa	2,500	8,741	2,170	1,090
Fungi	2,112	6,746	1,740	691
Bacteria	1,671	4,483	1,303	711
Eukaryota	478	2,867	765	338
Archaea	13	137	47	20
Unclassified	1,689	5,963	1,639	765

Supplementary Table S2. Distribution of NPs in each source kingdom or super-kingdom. Gradient red color indicates relative number of source species or NPs in each category. Green bars indicate percentage of active/potent NPs in each category (full-filled bar in the cell is 100%). *The same definition described in Table S1.

Data Source		No. of unique NPs	No. of unique NP-Species pairs
Manually curated from Publications		 24,927	 31,140
From Existing Databases	UNPD	 5,824	 183,346
	TCMID	 4,715	 15,782
	TM-MC	 3,693	 26,568
	TCMSP	 3,001	 14,530
	TCM_Taiwan	 2,597	 7,227
	HerDing	 1,537	 6,855
StreptomeDB		 798	 1,736

Supplementary Table S3. Data sources of source species of NPs. Manual inspection of publications contributed to source species annotation of most NPs. While the UNPD database contributed to most annotated NP-Species pairs. The source species of a specific NP may be annotated from multiple sources simultaneously. Green bars indicate percentage of NPs from each data source, while red bars show percentage of NP-Species pairs from each data source (full-filled bar in the cell is 100%).